

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ – ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ****Πρόγραμμα Μεταπτυχιακών Σπουδών  
«Κυβερνοασφάλεια και Επιστήμη Δεδομένων»****Μεταπτυχιακή Διατριβή**

Τίτλος Διατριβής	<b>Πρόβλεψη σοδειάς βιομηχανικής τομάτας με μεθόδους μηχανικής μάθησης.</b>  <b>Industrial tomato yield prediction using Machine Learning</b>
Όνοματεπώνυμο Φοιτητή	<b>ΚΑΣΙΜΑΤΗΣ ΧΡΙΣΤΟΦΟΡΟΣ-ΝΙΚΗΤΑΣ</b>
Πατρώνυμο	<b>ΑΝΤΩΝΙΟΣ</b>
Αριθμός Μητρώου	<b>ΜΠΚΕΔ21020</b>
Επιβλέπων	<b>Αποστόλου Δημήτριος, Καθηγητής</b>

Ημερομηνία Παράδοσης **Οκτώβριος 2023**

**Τριμελής Εξεταστική Επιτροπή**

Δημήτριος Αποστόλου  
Καθηγητής

Δρ. Γρηγόριος Κορωνάκος  
Διδάσκων ΠΜΣ

Δρ. Αλέξανδρος Μπουσδέκης  
Διδάσκων ΠΜΣ

## Περίληψη

Μοντέλα πρόβλεψης χρησιμοποιούνται όλο και περισσότερο λόγω των εξελίξεων στον τομέα της επιστήμης των υπολογιστών. Η ευρεία διαθεσιμότητα δεδομένων σε συνδυασμό με αλγόριθμους Μηχανικής Μάθησης επιτρέπει τον συσχετισμό μεταξύ μεταβλητών και την ακριβή πρόβλεψη της παραγωγής. Δεδομένα που προέρχονται από τη καλλιέργεια βιομηχανικής ντομάτας, η ντομάτα αποτελεί παγκοσμίως σε συχνότητα παραγωγής αλλά και κατανάλωσης λαχανικό, μπορούν να χρησιμοποιηθούν με σκοπό την ακριβή πρόβλεψη της παραγωγής. Σε αυτή τη μελέτη αναπτύχθηκε ένα μοντέλο Μηχανικής Μάθησης, το οποίο μπορεί να προβλέπει την παραγωγή της καλλιέργειας βιομηχανικής ντομάτας κατά τη διάρκεια της καλλιέργειάς της, βασιζόμενο σε δεδομένα που έχουν προηγουμένως συλλεχθεί (απόδοση, υβρίδιο που καλλιεργείται και κλιματικές συνθήκες της περιοχής). Αυτά τα δεδομένα, τα οποία προέρχονται από 302 διαφορετικά χωράφια σε 6 περιοχές της δυτικής Πελοποννήσου στην Ελλάδα, από το 2019 έως το 2021, δημιουργούν ένα σύνολο από διαφορετικά χαρακτηριστικά και πάνω από 600 καταγραφές. Για την εύρεση του βέλτιστου αλγορίθμου για το εν λόγω σύνολο δεδομένων, δοκιμάστηκαν διαφορετικοί αλγόριθμοι σε περισσότερες από 10 μετρήσεις. Το μοντέλο πρόβλεψης αναπτύχθηκε χρησιμοποιώντας τον αλγόριθμο Ridge. Το τελικό μοντέλο προέβλεψε την παραγωγή της ντομάτας εντός της περιόδου καλλιέργειας, ενημερώνοντας την πρόβλεψη μηνιαίως με τη χρήση νέων δεδομένων καιρού για 11 χωράφια κατά την καλλιεργητική περίοδο του 2022, με τη διαφορά μεταξύ των προβλεπόμενων και των πραγματικών τιμών να κυμαίνεται από μόλις 6 έως 6,674 κιλά.

## Abstract

Prediction models are extensively used due to improvements in computer science. The availability of data combined with machine learning algorithms allows the extraction of feature importance of the variables and accurate production forecasting. Cultivation data from industrial tomatoes, tomatoes are one of the most widely produced and consumed vegetables in the world, is used to accurately predict production. In this thesis, a model has been developed to predict the production of industrial tomato crops based on previously collected data. Data were collected from different fields in different regions of the Peloponnese in Greece, over 3 growing seasons. In order to find the optimal algorithm for this data set, many different algorithms were tested on different measurements. Ridge was used in order to develop the prediction model.

## Εισαγωγή

Οι τεχνολογικές εξελίξεις σήμερα έχουν αυξήσει τη χρήση των μοντέλων πρόβλεψης, καθώς είναι οικονομικότερα και παρέχουν πιο ακριβείς προβλέψεις (Ara et al., 2021). Τα μοντέλα αυτά δύνανται επίσης να προσφέρουν την υποστήριξη που μπορεί να χρειάζεται ένας/μία παραγωγός. Ο βασικός τους σκοπός είναι να πραγματοποιούν έγκαιρες και ακριβείς προβλέψεις, λαμβάνοντας υπόψη διάφορα δεδομένα που έχουν χρησιμοποιηθεί για την εκπαίδευσή τους και συσχετίζοντας μεταβλητές με έναν τρόπο που δεν μπορεί να εφαρμοστεί με τις παραδοσιακές στατιστικές μεθόδους. Υπάρχουν διάφορες εφαρμογές των μοντέλων πρόβλεψης στη γεωργία καθώς μπορούν να χρησιμοποιηθούν για τη μείωση των αναγκών σε άρδευση (Zimmermann και Fischer, 2020), τη μείωση των επιθέσεων από παράσιτα και ασθένειες που επηρεάζουν τις καλλιέργειες (Small et al., 2015), τη βελτίωση της ποιότητας των καλλιεργειών (Chawgien και Kiattisin, 2021) και την αύξηση της παραγωγής των καλλιεργειών (Malniya και Solanki, 2022). Επιπλέον, η ύπαρξη της δυνατότητας δημιουργίας διάφορων συνόλων δεδομένων όσον αφορά την ποικιλία και τον όγκο έχει δημιουργήσει την ανάγκη χρήσης διάφορων μοντέλων πρόβλεψης βάσει της διαθεσιμότητας δεδομένων του/της παραγωγού (Bouilla et al., 2018). Τα δεδομένα μπορούν να προέρχονται από αισθητήρες IoT (Psomakelis et al., 2016; Gupta και Nahar, 2022), από εικόνες δορυφόρων (M. Koller και S. K. Upadhyaya, 2005), από RGB εικόνες (Iraji, 2019) ή να είναι ακατέργαστα δεδομένα που περιέχουν πληροφορίες από προηγούμενα έτη, κλιματικά δεδομένα, μετρήσεις στον τόπο καλλιέργειας κ.λπ. (Nigam et al., 2019; Psomakelis et al., 2020b).

Η ντομάτα (*Solanum lycopersicum* L.) κατατάσσεται δεύτερη στην παραγωγή λαχανικών παγκοσμίως και είναι σημαντική για τη διατήρηση της ανθρώπινης υγείας (Leoni, 2003; Li et al., 2021). Υπάρχουν διάφορες καλλιέργειες ντομάτας, μία από τις οποίες είναι η βιομηχανική ντομάτα που επεξεργάζεται για την παραγωγή διάφορων προϊόντων. Τα τελευταία χρόνια έχουν υπάρξει πολλές προσεγγίσεις για τη βελτίωση της παραγωγής των καλλιεργειών στο πλαίσιο των Στόχων Βιώσιμης Ανάπτυξης των Ηνωμένων Εθνών καθώς και του Πράσινου Σχεδίου της Ευρωπαϊκής Ένωσης. Αυτό μπορεί να επιτευχθεί με τη χρήση περιβαλλοντικά φιλικών τεχνικών (Katsenios et al., 2022) ή με τη χρήση τεχνολογικών μέσων (Lázaro et al., 2021).

Ο στόχος αυτής της μελέτης ήταν η ανάπτυξη ενός μοντέλου που να είναι σε θέση να προβλέπει την παραγωγή της βιομηχανικής ντομάτας κατά τη διάρκεια της καλλιέργειας, χρησιμοποιώντας ως καταχωρήσεις δεδομένα από μετεωρολογικούς σταθμούς, προηγούμενη παραγωγή και το υβρίδιο που καλλιεργήθηκε. Επιπλέον, ήταν σημαντικό να εντοπιστούν οι κλιματικές συνθήκες που επηρεάζουν την παραγωγή της ντομάτας καθώς η Ελλάδα χαρακτηρίζεται από ζεστά και ξηρά καλοκαίρια. Έτσι, εξήχθη η συσχέτιση μεταξύ των μετεωρολογικών μεταβλητών, του είδους ντομάτας που καλλιεργήθηκε και της παραγωγής. Τέλος, το μοντέλο ενημερώνει την πρόβλεψη κάθε φορά που προστίθενται νέα μετεωρολογικά δεδομένα, γεγονός που το καθιστά πιο ακριβές.

## Σχετική έρευνα

Τα μοντέλα πρόβλεψης χρησιμοποιούνται επί του παρόντος σε διάφορες καλλιέργειες. Τα κλιματικά δεδομένα, η παραγωγή καθώς και ο τύπος της καλλιέργειας έχουν χρησιμοποιηθεί ως δεδομένα πρόβλεψης για κατηγοριοποίηση των αγρών με βάση την αναμενόμενη παραγωγή (Gandhi et al., 2016). Σε έρευνα που αφορά το βαμβάκι χρησιμοποιήθηκαν δεδομένα εδάφους και παραγωγής για την πρόβλεψη της παραγωγής και τη βελτίωση της διαχείρισης της καλλιέργειας (Parageorgiou et al., 2011). Σε άλλη έρευνα για το καλαμπόκι και τη σόγια προτάθηκε για την πρόβλεψη της παραγωγής της καλλιέργειας μοντέλο που χρησιμοποιεί δεδομένα από δορυφόρους και λαμβάνει υπόψη και τη μέση παραγωγή (Khaki et al., 2021). Σε μια άλλη μελέτη αναπτύχθηκε ένα μοντέλο βασισμένο σε ασαφείς λογικές, το οποίο χρησιμοποιεί δεδομένα εδάφους, καιρού, ύδρευσης και λίπανσης προσπαθώντας να αυξήσει την παραγωγή και να μειώσει τη χρήση λιπασμάτων (Prabakaran et al., 2018). Σε μια μελέτη για παράσιτα, δοκιμάστηκε ένα μοντέλο μηχανικής μάθησης (ML) με σκοπό να βρει τον πληθυσμό των επιβλαβών εντόμων σε φυτώρια, βοηθώντας τους χρήστες να προγραμματίσουν τον ψεκασμό εκ των προτέρων (Rupnik et al., 2019).

Τα τελευταία χρόνια, έχουν αναπτυχθεί μοντέλα πρόβλεψης για τις καλλιέργειες ντοματών σε θερμοκήπια. Σε μια πρόσφατη μελέτη, αναπτύχθηκε ένα μοντέλο μηχανικής μάθησης με τη χρήση διαφόρων προσεγγίσεων συγχώνευσης για την πρόβλεψη της παραγωγής σε θερμοκήπια ντομάτας. Το σύνολο δεδομένων περιλάμβανε δεδομένα από αισθητήρες IoT (Gong et al., 2022). Σε μια άλλη έρευνα σχετικά με την καλλιέργεια ντοματών σε θερμοκήπια, χρησιμοποιήθηκε ένα Νευρωνικό Δίκτυο με Γενετικό Αλγόριθμο (Neural Network with Genetic Algorithm) (Wang et al., 2021). Επιπλέον, χρησιμοποιήθηκαν δεδομένα από προηγούμενες παραγωγές και περιβαλλοντικές συνθήκες στα θερμοκήπια για την ανάπτυξη ενός μοντέλου βασισμένου σε Νευρωνικά Δίκτυα (NNs) με σκοπό τη βελτίωση της παραγωγικότητας των καλλιεργειών μέσω της βελτίωσης των συνθηκών στα θερμοκήπια (Gong et al., 2021). Παρόμοιες έρευνες διεξήχθησαν στη Γερμανία, χρησιμοποιώντας παρελθοντικά κλιματικά δεδομένα ως εισαγωγή για την ανάπτυξη ενός Νευρωνικού Δικτύου (NN) για την πρόβλεψη της παραγωγής σε καλλιέργειες ντομάτας σε θερμοκήπια με εβδομαδιαία βάση (Salazar et al., 2015, 2017). Ένα άλλο μοντέλο αναπτύχθηκε χρησιμοποιώντας Τεχνητά Νευρωνικά Δίκτυα (ANNs) σε ένα σύνολο δεδομένων που περιλάμβανε πληροφορίες σχετικά με τα λιπάσματα, τις ανάγκες σε νερό, την κατανάλωση ενέργειας, προσπαθώντας έτσι να προβλέψει την παραγωγή ντομάτας (Belouz et al., 2022).

Η πρόβλεψη της παραγωγής στα χωράφια κατά τη διάρκεια της καλλιέργειας μπορεί να επιτευχθεί χρησιμοποιώντας διάφορες τεχνικές. Η ανάπτυξη ενός μοντέλου πρόβλεψης για καλλιέργειες κατά τη διάρκεια της ανάπτυξης τους στο χωράφι είναι διαφορετική διαδικασία σε σύγκριση με τις καλλιέργειες σε θερμοκήπια, δεδομένου ότι οι κλιματικές συνθήκες είναι απρόβλεπτες και δεν μπορούν να τυποποιηθούν. Ωστόσο, ένα πλεονέκτημα των καλλιεργειών στο χωράφι είναι η δυνατότητα χρήσης δεδομένων από δορυφόρους για την εξαγωγή χρήσιμων δεικτών (π.χ., NDVI, LAI, NDWI κ.λπ.). Σε μια πρόσφατη μελέτη, εξήχθησαν 6 Δείκτες Φυτικότητας (Vegetation Indices - VIs) με σκοπό την πρόβλεψη της παραγωγής της καλλιέργειας (Psiroukis et al., 2022). Επίσης, τα μη επανδρωμένα εναέρια μέσα (Unmanned Aerial Vehicles - UAVs) μπορούν να χρησιμοποιηθούν για την εξαγωγή των VIs. Διαφορετικοί αλγόριθμοι δοκιμάστηκαν σε μια μελέτη σχετική με την πρόβλεψη παραγωγής σε καλλιέργεια ντομάτας σε μία προσπάθεια να εξαχθεί η βιομάζα των φρέσκων φυτών αλλά και το βάρος καθώς και ο αριθμός των καρπών. Κάθε μία από τις εν λόγω μεταβλητές προβλεπόταν χρησιμοποιώντας τον αλγόριθμο που είχε την καλύτερη ακρίβεια στην πρόβλεψή της (Tatsumi et al., 2021). Εικόνες των καρπών ντομάτας χρησιμοποιήθηκαν για την εκπαίδευση διαφόρων Νευρωνικών Δικτύων (NNs) (VGG19, YOLOv3) με σκοπό την μέτρηση του αριθμού των καρπών και την εκτίμηση της παραγωγής (Maheswari et al., 2022; Wang et al., 2022). Σε μια παρόμοια έρευνα, χρησιμοποιήθηκαν συνελκτικά NNs για τον εντοπισμό του αριθμού των καρπών σε κάθε δέντρο και την εκτίμηση της παραγωγής του (Liu, 2018). Ένα μοντέλο που χρησιμοποιεί δεδομένα εδάφους, πληροφορίες για την καλλιέργεια και τη λίπανση αναπτύχθηκε με σκοπό τη βελτιστοποίηση της διαχείρισης του ύδατος για το πότισμα καθώς και την πρόβλεψη της παραγωγής (Dehghanisanij et al., 2021).

## Μεθοδολογία Έρευνας

Για την ανάπτυξη ενός μοντέλου πρόβλεψης της παραγωγής βιομηχανικής ντομάτας κατά τη διάρκεια της καλλιέργειας, συλλέχθηκαν δεδομένα από 302 διαφορετικά χωράφια σε έξι περιοχές της δυτικής Πελοποννήσου στην Ελλάδα, κατά τη διάρκεια τριών περιόδων καλλιέργειας (2019-2021). Το σύνολο δεδομένων που χρησιμοποιήθηκε για την ανάπτυξη του μοντέλου είχε τρεις διαφορετικές πηγές δεδομένων. Συγκεκριμένα, χρησιμοποιήθηκαν δεδομένα κλιματικών συνθηκών από την περιοχή, δεδομένα εδάφους από τα χωράφια και, τέλος, δεδομένα σχετικά με τις καλλιέργειες (παραγωγή και υβρίδιο ντομάτας) που χρησιμοποιήθηκαν. Τα δεδομένα σχετικά με την παραγωγή και τα υβρίδια ντομάτας που καλλιεργήθηκαν παρασχέθηκαν από την εταιρεία κονσερβοποιίας ΚΥΚΝΟΣ S.A. στην Ελλάδα. Μετά την ανάπτυξη του μοντέλου πρόβλεψης της παραγωγής, για να αξιολογηθεί, επιλέχθηκαν 11 διαφορετικά χωράφια στην πειραματική περιοχή για την καλλιεργητική περίοδο του 2022.

## Τεχνολογίες που χρησιμοποιήθηκαν

Χρησιμοποιήσαμε τη βάση δεδομένων MongoDB για την αποθήκευση των δεδομένων μας. Αυτή η βάση δεδομένων μας προσέφερε ορισμένα πλεονεκτήματα που μας επέτρεψαν να χειριστούμε τα δεδομένα όπως επιθυμούσαμε. Αρχικά η MongoDB είναι βάση υψηλής επεκτασιμότητας, πράγμα που μας επέτρεψε να διευρύνουμε το σύνολο προσθέτοντας περισσότερα δεδομένα όταν απαιτούνταν. Επιπλέον, είναι μια βάση δεδομένων NoSQL που χρησιμοποιεί τη μορφή JSON. Αυτό σημαίνει ότι η μορφή των δεδομένων μπορεί να τροποποιηθεί ανάλογα με τις ανάγκες μας, επιτρέποντάς μας να προσθέτουμε νέες μετρήσεις και πεδία χωρίς σημαντικές τροποποιήσεις στα υπάρχοντα δεδομένα (Psomakelis et al., 2020a; Makris et al., 2022). Η μηχανική μάθηση πραγματοποιήθηκε χρησιμοποιώντας τη γλώσσα προγραμματισμού Python 3.9, καθώς και τη βιβλιοθήκη Scikit Learn (Buitinck et al., 2013; Pedregosa et al., 2011).

## Σύνολο δεδομένων

Για την εξέταση της παραγωγής, χρησιμοποιήσαμε ένα σύνολο δεδομένων που αποτελείται από 636 εγγραφές παραγωγής ντομάτας από 302 διαφορετικά χωράφια για 3 συνεχόμενα έτη, από το 2019 έως το 2021. Αυτές οι εγγραφές περιλαμβάνουν 31 διαφορετικά χαρακτηριστικά, συμπεριλαμβανομένων δεδομένα εδάφους από το χωράφι, μετρήσεις καιρού ανά μήνα, τον τύπο υβριδίου ντομάτας που χρησιμοποιήθηκε καθώς και την παραγωγή του χωραφιού για κάθε εποχή. Ειδικότερα, οι θερμοκρασίες (μέση, μέγιστη, ελάχιστη) για κάθε μήνα μετρήθηκαν σε °C, το ύψος των βροχοπτώσεων κάθε μήνα μετρήθηκε σε mm (χιλιοστόμετρα). Ο τύπος του χωραφιού, τα ποσοστά αργίλου, ιλύς και άμμου στο χωράφι, μετρήθηκαν χρησιμοποιώντας τη μέθοδο του Βουγουκος (1962) και την ταξινόμηση εδάφους του USDA (1999). Τα χαρακτηριστικά του εδάφους μετρήθηκαν μια φορά, δεδομένου ότι η υφή του εδάφους του χωραφιού είναι ένα φυσικό χαρακτηριστικό που παραμένει ίδιο για τα τρία έτη των μετρήσεων.

## Κλιματολογικά δεδομένα

Τα κλιματολογικά δεδομένα που χρησιμοποιήθηκαν για την ανάπτυξη του μοντέλου πρόβλεψης εξήχθησαν από μετεωρολογικούς σταθμούς του Εθνικού Αστεροσκοπείου Αθηνών, οι οποίοι συγκεντρώνουν δεδομένα (θερμοκρασία, βροχή, υγρασία κ.λπ.) κάθε 10 λεπτά (Lagouvardos et al., 2017). Τα μετεωρολογικά δεδομένα που χρησιμοποιήθηκαν περιλάμβαναν τη μέση θερμοκρασία, τη μέγιστη θερμοκρασία, την ελάχιστη θερμοκρασία και την ποσότητα βροχής για κάθε μήνα από Απρίλιο έως Ιούλιο. Ο λόγος που δεν χρησιμοποιήθηκαν μετεωρολογικά δεδομένα για ολόκληρο το έτος είναι ότι οι ντομάτες καλλιεργούνται τον Απρίλιο και συλλέγονται από τον Ιούλιο μέχρι τον Αύγουστο, επομένως περισσότερα δεδομένα θα έκαναν μόνο το μοντέλο πιο αργό και πολύπλοκο. Επιπλέον, τα μετεωρολογικά δεδομένα χρησιμοποιούνται για την εξαγωγή συσχετίσεων μεταξύ της παραγωγής και των κλιματικών συνθηκών.

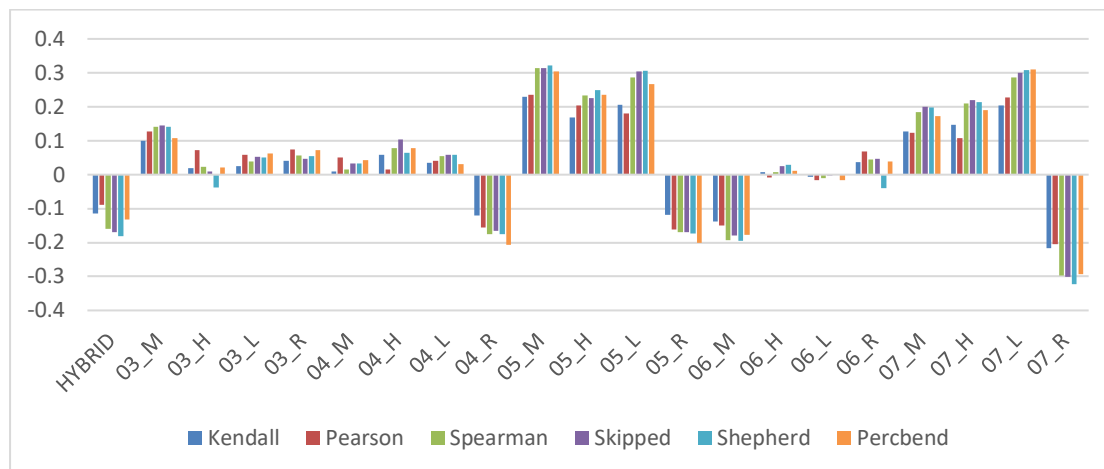
## Υβρίδια Ντομάτας

Συνολικά, 30 διαφορετικά είδη υβριδίων ντομάτας χρησιμοποιήθηκαν μέσα στα 3 χρόνια των πειραμάτων. Το υβρίδιο που χρησιμοποιήθηκε περισσότερο είναι το HEINZ1015 το οποίο εμφανίζεται συνολικά 227 φορές.

## Στατιστική Ανάλυση του συνόλου δεδομένων

Για να αξιολογήσουμε και να κατανοήσουμε τις μεταβλητές στο σύνολο δεδομένων μας, εξαγάγαμε διάφορες στατιστικές συσχετίσεις για τη μεταβλητή της παραγωγής. Αυτό μας βοήθησε να βρούμε τις

μεταβλητές που σχετίζονται με την παραγωγή και πόσο σημαντικές είναι. Στο Σχήμα 1 μπορούμε να δούμε την συνδιακύμανση που εξήχθη με τη χρήση 6 διαφορετικών στατιστικών μεθόδων.



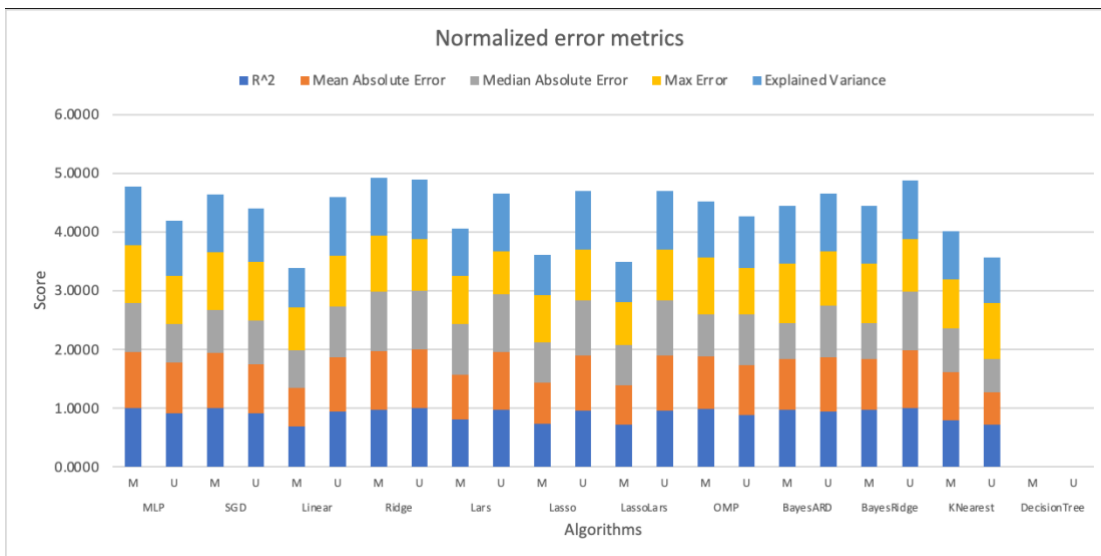
**Σχήμα 1.** Συνδιακύμανση παραγωγής χρησιμοποιώντας 6 στατιστικές μεθόδους. 03\_M: Μέση θερμοκρασία του Μαρτίου. 03\_H: Μέγιστη θερμοκρασία του Μαρτίου. 03\_L: Ελάχιστη θερμοκρασία του Μαρτίου. 03\_R: Βροχόπτωση του Μαρτίου.

Το Σχήμα 1 δείχνει ότι οι 6 μέθοδοι που χρησιμοποιήθηκαν έχουν παρόμοια αποτελέσματα. Σε όλες τις χρησιμοποιούμενες μεθόδους, οι πιο σημαντικές μεταβλητές φαίνεται να είναι το κλίμα του Μαΐου (μέση, υψηλή και χαμηλή θερμοκρασία), καθώς και η βροχόπτωση και η χαμηλή θερμοκρασία του Ιουλίου. Τέλος, σημαντική είναι επίσης η βροχόπτωση του Μαΐου και του Απριλίου.

## Μοντέλα Μηχανικής Μάθησης

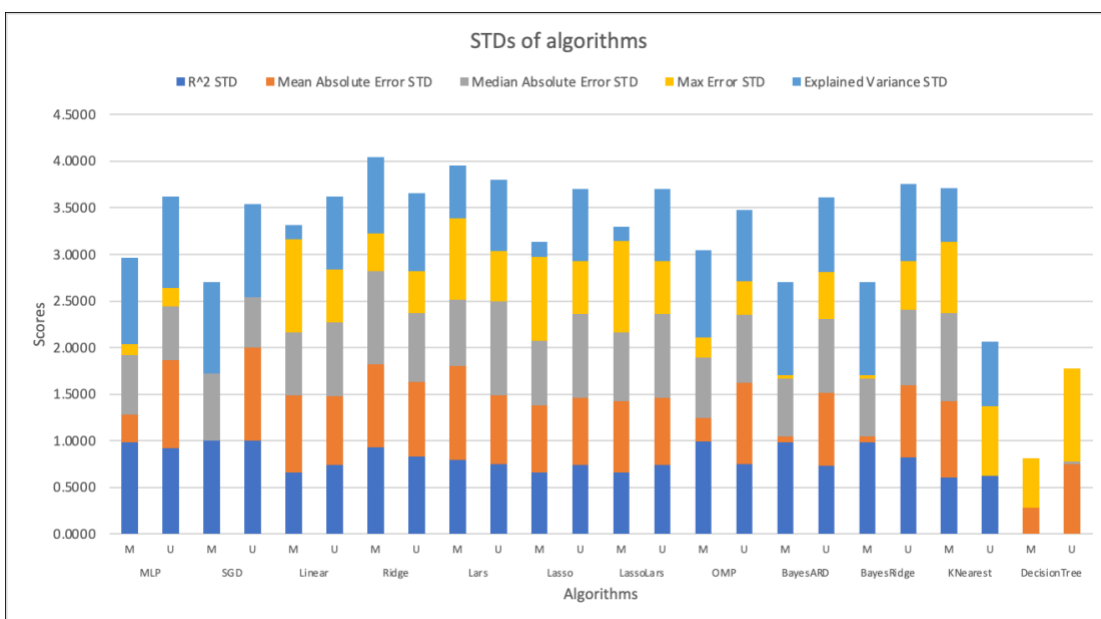
Η πρώτη διαδικασία που πραγματοποιήσαμε ήταν η εξαγωγή ορισμένων στατιστικών χαρακτηριστικών προκειμένου να κατανοήσουμε το σύνολο δεδομένων που θα χρησιμοποιούσαμε για την εκπαίδευση του μοντέλου, τόσο με τα δεδομένα του εδάφους όσο και χωρίς αυτά. Αυτό σημαίνει ότι τα συγκεκριμένα χαρακτηριστικά εξήχθησαν δύο φορές, μία με τα δεδομένα εδάφους και μία χωρίς αυτά. Το σύνολο δεδομένων που χρησιμοποιήσαμε περιλάμβανε πληροφορίες για 302 διάφορα χωράφια, ενώ μόνο 54 από αυτά περιείχαν δεδομένα εδάφους, με αποτέλεσμα το 25,97% των χωραφιών να περιλαμβάνουν και τα δύο είδη δεδομένων. Επιπλέον, αυτά τα 54 χωράφια που περιείχαν δεδομένα εδάφους εμφανίστηκαν 126 φορές στο αρχικό σύνολο δεδομένων. Το γεγονός ότι το αρχικό σύνολο δεδομένων περιλαμβάνει 636 δείγματα χωραφιών, σημαίνει πως το 19,81% των συνολικών εγγραφών στη βάση δεδομένων περιέχουν και τα δύο είδη δεδομένων.

Δώδεκα διαφορετικοί αλγόριθμοι δοκιμάστηκαν με 10-fold cross validation σε μια προσπάθεια να βρεθεί η τιμή πρόβλεψης του συνόλου δεδομένων που χρησιμοποιήθηκε. Συνολικά, εκπαιδεύτηκαν περισσότερα από 20.000 διαφορετικά μοντέλα πριν από τη συγχώνευση των δύο διαφορετικών συνόλων δεδομένων αλλά και μετά από αυτή, σε μια προσπάθεια να βελτιστοποιηθεί ο αλγόριθμος που είχε τα καλύτερα αποτελέσματα σε συνδυασμό με τις υπερπαραμέτρους (hyper-parameters) του. Τα σύνολα δεδομένων παρέμειναν αναλλοίωτα και συγχωνεύτηκαν μόνο για τις ανάγκες ορισμένων πειραμάτων. Μια μελέτη που συγκρίνει τα σκορ των αλγορίθμων μπορεί να βρεθεί στο Σχήμα 2.



**Σχήμα 2.** Σύγκριση των μετρήσεων σφαλμάτων που εξήχθησαν για όλους τους αλγόριθμους (μετά την κανονικοποίηση) με και χωρίς τα δεδομένα εδάφους. M: σκορ των μοντέλων που εκπαιδεύτηκαν με το σύνολο δεδομένων που περιέχει δεδομένα εδάφους, U: σκορ των μοντέλων που εκπαιδεύτηκαν με το σύνολο δεδομένων που δεν περιέχει δεδομένα εδάφους.

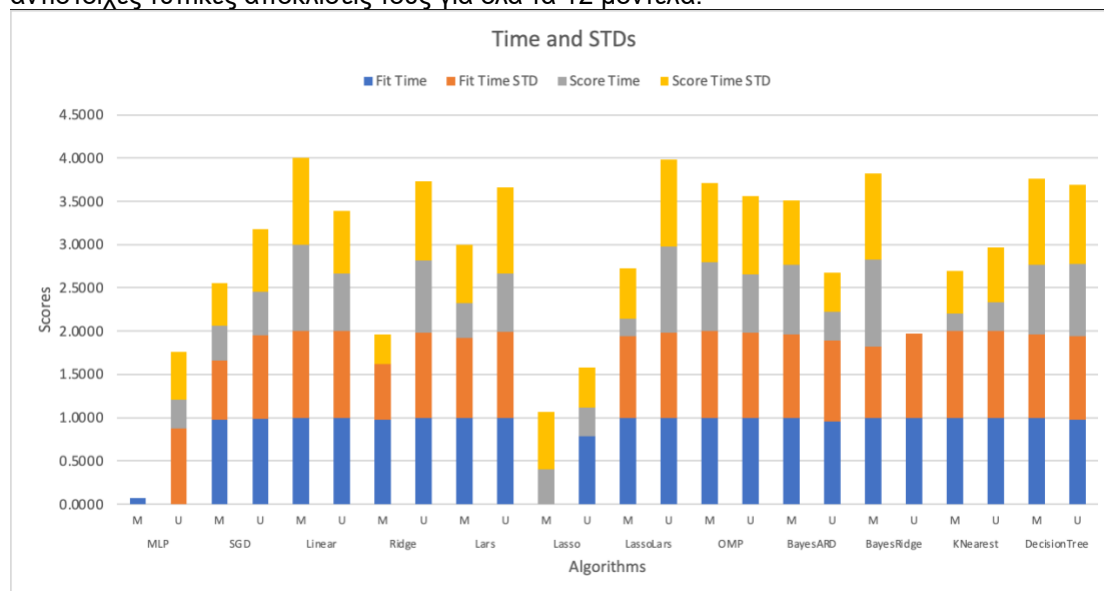
Όπως απεικονίζεται, πέντε διαφορετικές μετρικές σφάλματος έχουν υπολογιστεί για κάθε έναν από τα 12 αλγόριθμους, έχοντας κανονικοποιηθεί στο εύρος 0-1. Είναι σαφές ότι το μοντέλο Ridge είναι αποδοτικότερο από κάθε άλλον αλγόριθμο, τόσο με όσο και χωρίς τη χρήση των δεδομένων εδάφους. Επιπλέον, έχει την καλύτερη μετρική  $R^2$  και το υψηλότερο μέσο σφάλμα. Αυτό σημαίνει ότι το συγκεκριμένο σύνολο δεδομένων ταιριάζει καλά σε αυτό το μοντέλο και κατανοεί τις τιμές που είναι έκτοπες. Στο Σχήμα 3 συγκρίνονται οι τυπικές αποκλίσεις (STDs) για τις 5 μετρικές για τους 12 δοκιμασμένους αλγόριθμους. Οι εξαγμένες STDs επιτρέπουν να αξιολογηθεί και να εκτιμηθεί η αξιοπιστία της απόδοσης αυτών των 12 διαφορετικών αλγόριθμων. Αυτές οι μετρικές έχουν επίσης κανονικοποιηθεί στο εύρος 0-1. Όσο πιο κοντά στο 1 είναι η τιμή τους, τόσο πιο σταθερό είναι το μοντέλο. Ελέγχοντας το σχήμα, βλέπουμε ότι οι αλγόριθμοι Ridge και Lars είναι οι πιο σταθεροί. Επιπλέον, ο αλγόριθμος Ridge έχει υψηλή STD στο Max Error, ενώ ο αλγόριθμος Lars έχει υψηλότερη STD στο Explained Variance και Mean Absolute Error.





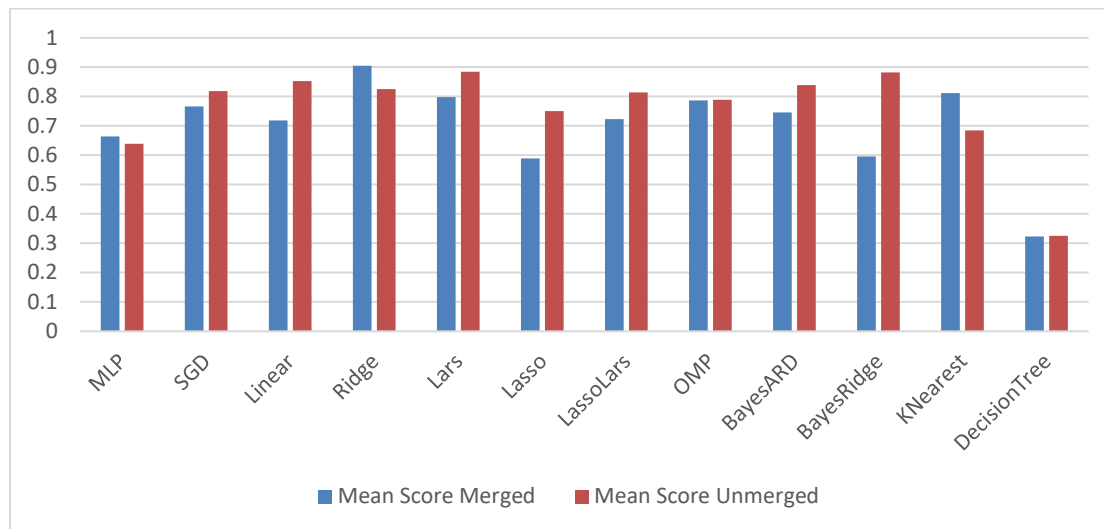
**Σχήμα 3.** Σύγκριση της κανονικοποιημένης τυπικής απόκλισης των μετρικών σφαλμάτων για όλους τους αλγόριθμους (μετά την κανονικοποίηση) με και χωρίς τα δεδομένα εδάφους. M: σκορ των μοντέλων που εκπαιδεύτηκαν με το σύνολο δεδομένων που περιέχει δεδομένα εδάφους, U: σκορ των μοντέλων που εκπαιδεύτηκαν με το σύνολο δεδομένων που δεν περιέχει δεδομένα εδάφους.

Τέλος, είναι σημαντικό να γνωρίζουμε πόσο γρήγορα μπορεί να εκπαιδευτεί και να αξιολογηθεί ένα μοντέλο. Το Σχήμα 4 απεικονίζει δύο μεταβλητές σχετικά με τον χρόνο που απαιτείται, καθώς και τις αντίστοιχες τυπικές αποκλίσεις τους για όλα τα 12 μοντέλα.



**Σχήμα 4.** Σύγκριση του απαιτούμενου χρόνου και των αντίστοιχων τυπικών αποκλίσεων που εξήχθησαν για όλους τους αλγόριθμους (μετά την κανονικοποίηση) με και χωρίς τα δεδομένα εδάφους. M: σκορ των μοντέλων που εκπαιδεύτηκαν με το σύνολο δεδομένων που περιέχει δεδομένα εδάφους, U: σκορ των μοντέλων που εκπαιδεύτηκαν με το σύνολο δεδομένων που δεν περιέχει δεδομένα εδάφους.

Το Σχήμα 4 απεικονίζει τον μέσο χρόνο που απαιτείται για τη βαθμολόγηση και την προσαρμογή ενός μοντέλου. Επιπλέον, εξήχθησαν οι τυπικές αποκλίσεις (STDs) για αυτές τις μετρικές. Και πάλι, αυτές οι μετρικές έχουν κανονικοποιηθεί στο εύρος 0-1. Ο αλγόριθμος LassoLars αποδίδει καλύτερα από οποιονδήποτε άλλον αλγόριθμο στο σύνολο δεδομένων χωρίς δεδομένα εδάφους, ενώ ο αλγόριθμος Linear αποδίδει καλύτερα όταν χρησιμοποιούνται δεδομένα εδάφους. Οι αλγόριθμοι Ridge, Lars και Decision Trees επίσης αποδίδουν καλά χωρίς δεδομένα εδάφους, αλλά έχουν χαμηλότερο χρόνο βαθμολόγησης από τον LassoLars. Όλα αυτά μας βοήθησαν να επιλέξουμε τον αλγόριθμο Ridge για την εκπαίδευση του μοντέλου. Επιπλέον, ο Ridge έχει το καλύτερο μέσο σκορ σε 14 μετρικές που χρησιμοποιήθηκαν, όταν χρησιμοποιούνται δεδομένα εδάφους. Όταν δεν χρησιμοποιούνται δεδομένα εδάφους, ο Ridge είχε το 5ο καλύτερο μέσο σκορ από όλες τις μετρικές. Ωστόσο, είναι σημαντικό να δοθεί προτεραιότητα σε αυτό που χρειαζόμαστε για το πείραμά μας. Προτιμήσαμε να χρησιμοποιήσουμε έναν αλγόριθμο που αποδίδει καλά ακόμη κι αν δεν είναι ο γρηγορότερος. Το μέσο σκορ κάθε δοκιμασμένου αλγορίθμου απεικονίζεται στο Σχήμα 5.

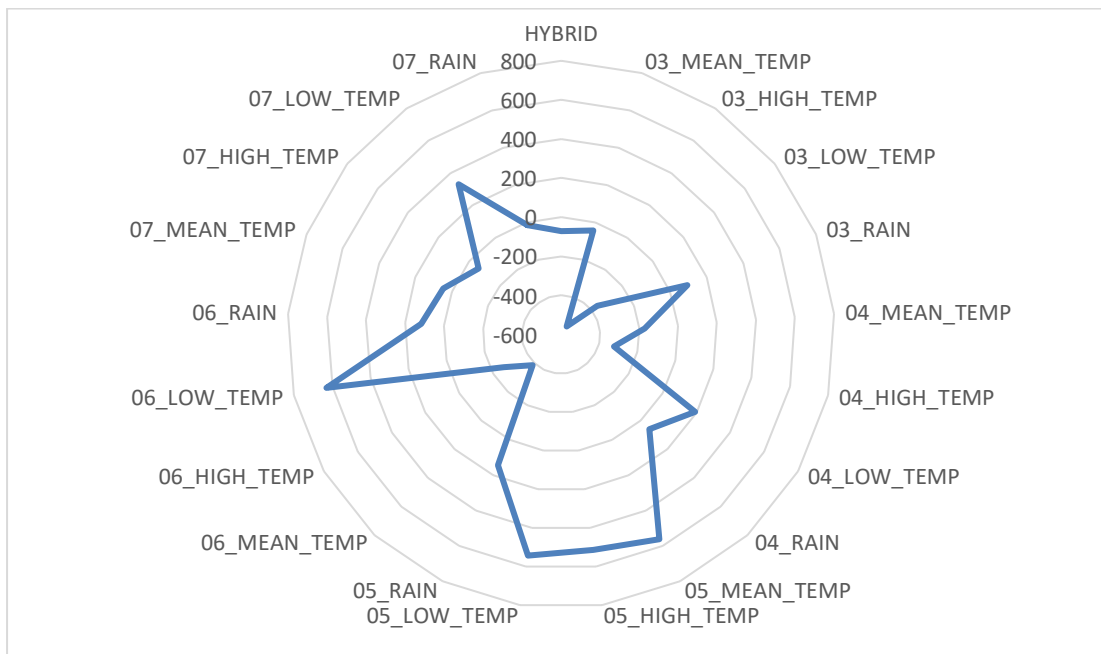


Σχήμα 5. Μέσο σκορ όλων των μετρικών για κάθε αλγόριθμο (μετά την κανονικοποίηση).

## Αποτελέσματα

Για την ανάπτυξη του μοντέλου πρόβλεψης σοδιάς, ξεκινήσαμε την εκπαίδευση του μοντέλου χρησιμοποιώντας δύο διαφορετικά σύνολα δεδομένων. Η κύρια διαφορά μεταξύ τους ήταν η ύπαρξη ή όχι δεδομένων εδάφους. Για να καταλήξουμε στο σύνολο δεδομένων που θα χρησιμοποιήσουμε για την εκπαίδευση του τελικού μοντέλου, συγκρίναμε τη μετρική  $R^2$  πριν από την κανονικοποίησή της, για τον βέλτιστο αλγόριθμο σε κάθε σύνολο δεδομένων. Το σύνολο δεδομένων που περιείχε δεδομένα εδάφους είχε τιμή  $R^2$  ίση με 0,09, ενώ το σύνολο δεδομένων χωρίς δεδομένα εδάφους είχε τιμή  $R^2$  ίση με 0,144. Ως αποτέλεσμα, αποφασίσαμε να εκπαιδεύσουμε το τελικό μοντέλο με το σύνολο που δεν περιείχε δεδομένα εδάφους. Μια εξήγηση για το ότι τα δεδομένα εδάφους δεν προσθέτουν περαιτέρω αξία στο μοντέλο είναι ότι μόνο το 26% των συνολικών δεδομένων περιείχε δεδομένα εδάφους. Αυτός ήταν ένας ακόμη λόγος για τη μην χρησιμοποιηθούν τα δεδομένα εδάφους, καθώς το σύνολο δεδομένων εκπαίδευσης του μοντέλου θα ήταν σημαντικά μικρότερο.

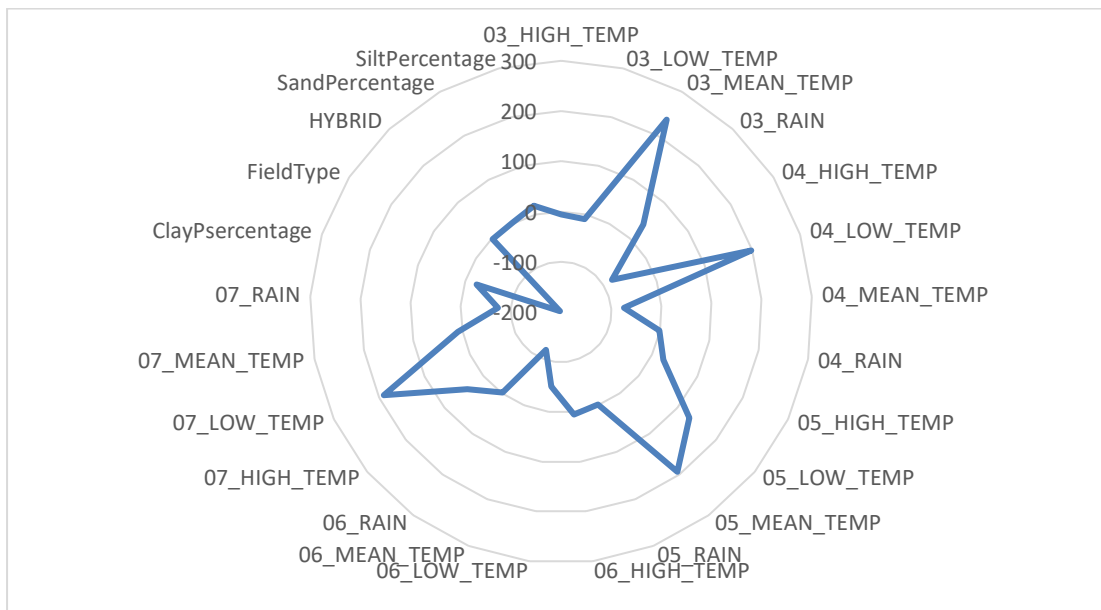
Η πρόβλεψη της απόδοσης κατά τη διάρκεια ή ακόμη και πριν την περίοδο καλλιέργειας είναι σημαντική τόσο για τους αγρότες όσο και για τις βιομηχανίες που επεξεργάζονται τις ντομάτες. Για να βρούμε τις μεταβλητές και τον τρόπο που επηρεάζουν το μοντέλο, εξαγάγαμε τη συσχέτιση των χαρακτηριστικών του συνόλου δεδομένων χωρίς δεδομένα εδάφους, χρησιμοποιώντας τον αλγόριθμο Ridge. Τα αποτελέσματα έδειξαν ότι η ελάχιστη θερμοκρασία των μηνών Ιουνίου και Μαΐου είναι οι πιο σημαντικές μεταβλητές όσον αφορά την απόδοση της βιομηχανικής ντομάτας. Επιπλέον, η μέγιστη θερμοκρασία των μηνών Μαΐου και Ιουλίου διαδραματίζει επίσης σημαντικό ρόλο. Το Σχήμα 6 απεικονίζει τη συσχέτιση των μεταβλητών χρησιμοποιώντας τον αλγόριθμο Ridge.



**Σχήμα 6. Συσχέτιση χαρακτηριστικών χρησιμοποιώντας τον αλγόριθμο Ridge στο σύνολο δεδομένων που δεν περιείχε δεδομένα εδάφους. 03\_M: Μέση θερμοκρασία του Μαρτίου; 03\_H: Μέγιστη θερμοκρασία του Μαρτίου; 03\_L: Ελάχιστη θερμοκρασία του Μαρτίου; 03\_R: Βροχόπτωση του Μαρτίου.**

Επιπλέον, ορισμένες από τις μεταβλητές φαίνεται να έχουν παρόμοια συσχέτιση, όταν εξάγονται με τη χρήση παραδοσιακών στατιστικών μεθόδων, ωστόσο, υπάρχουν σημαντικές διαφορές. Αυτός είναι ο λόγος που εξαγάγαμε τη συσχέτιση των μεταβλητών χρησιμοποιώντας αλγόριθμους μηχανικής μάθησης, καθώς κατανοούν τα δεδομένα με διαφορετικό τρόπο και κάνουν ακριβείς προβλέψεις. Ακολουθήσαμε την ίδια διαδικασία για το σύνολο δεδομένων που περιλάμβανε δεδομένα εδάφους, προκειμένου να δούμε αν η προσθήκη τους αλλάζει τη συσχέτιση των μεταβλητών. Επιπλέον, ήταν σημαντικό να βρούμε αν υπήρχε οποιαδήποτε ομοιότητα μεταξύ της συσχέτισης των χαρακτηριστικών (feature coefficient) με και χωρίς εδαφικά δεδομένα. Ο αλγόριθμος που δοκιμάστηκε ήταν και πάλι ο Ridge, αφού είχε την καλύτερη απόλυτη μετρική  $R^2$ .

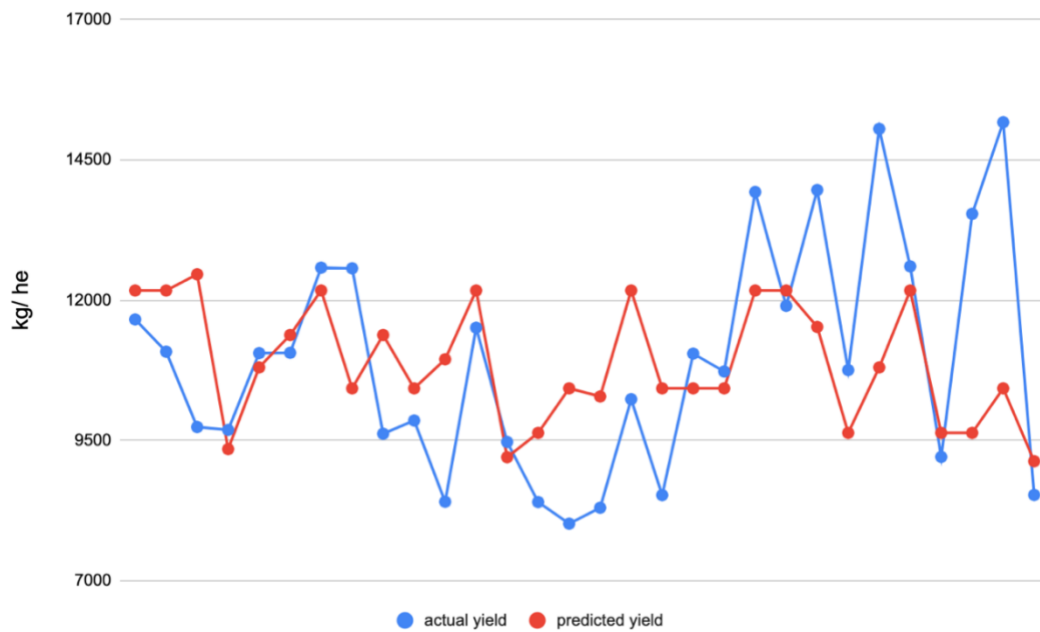
Η συσχέτιση των μεταβλητών του συνόλου δεδομένων που περιείχε δεδομένα εδάφους έδειξε ότι η ελάχιστη θερμοκρασία του Ιουλίου και του Απριλίου, καθώς και η μέση θερμοκρασία του Μαΐου και του Μαρτίου έχουν θετική επίδραση στην παραγωγή, ενώ το ποσοστό του αργίλου του εδάφους, η μέση θερμοκρασία του Ιουνίου και του Απριλίου, καθώς και η μέγιστη θερμοκρασία του Απριλίου είναι αντιστρόφως ανάλογες με την απόδοση. Το Σχήμα 7 απεικονίζει τη συσχέτιση των χαρακτηριστικών χρησιμοποιώντας το μοντέλο Ridge στο σύνολο δεδομένων που περιέχει τα δεδομένα εδάφους.



**Σχήμα 7. Συσχέτιση χαρακτηριστικών χρησιμοποιώντας τον αλγόριθμο Ridge σε ένα σύνολο δεδομένων που περιλαμβάνει δεδομένα για το έδαφος. 03\_M: Μέση θερμοκρασία του Μαρτίου; 03\_H: Μέγιστη θερμοκρασία του Μαρτίου; 03\_L: Ελάχιστη θερμοκρασία του Μαρτίου; 03\_R: Βροχοπτώσεις του Μαρτίου.**

Εφαρμόσαμε μοντέλα μηχανικής μάθησης προκειμένου να προβλέψουμε την παραγωγή της βιομηχανικής τομάτας κατά τη διάρκεια της καλλιέργειας. Αυτό επιτεύχθηκε δοκιμάζοντας το μοντέλο πρόβλεψης που αναπτύξαμε σε ένα σύνολο υποθετικών σεναρίων, προσπαθώντας να προβλέψουμε την τελική παραγωγή κάθε πεδίου κατά τη διάρκεια της καλλιέργειας. Αυτό είναι δυνατό να επιτευχθεί λόγω της ικανότητας των μοντέλων μηχανικής μάθησης να δημιουργούν σχέσεις μεταξύ της μεταβλητής-στόχου και των διαθέσιμων δεδομένων.

Τα αποτελέσματα του μοντέλου έδειξαν ότι η πρόβλεψη της παραγωγής στο σύνολο δεδομένων μας ήταν κοντά στην πραγματική παραγωγή. Όταν προστίθενται νέα δεδομένα, η πρόβλεψη της παραγωγής τους θα βασίζεται στα δεδομένα πάνω στα οποία εκπαιδεύτηκε το μοντέλο, προσπαθώντας να εντοπίσει ομοιότητες. Το Σχήμα 8 απεικονίζει το διάγραμμα πρόβλεψης για 30 διαφορετικά πεδία που επιλέχθηκαν τυχαία. Όπως μπορούμε να δούμε, η προβλεπόμενη παραγωγή είναι κοντά στην πραγματική παραγωγή.

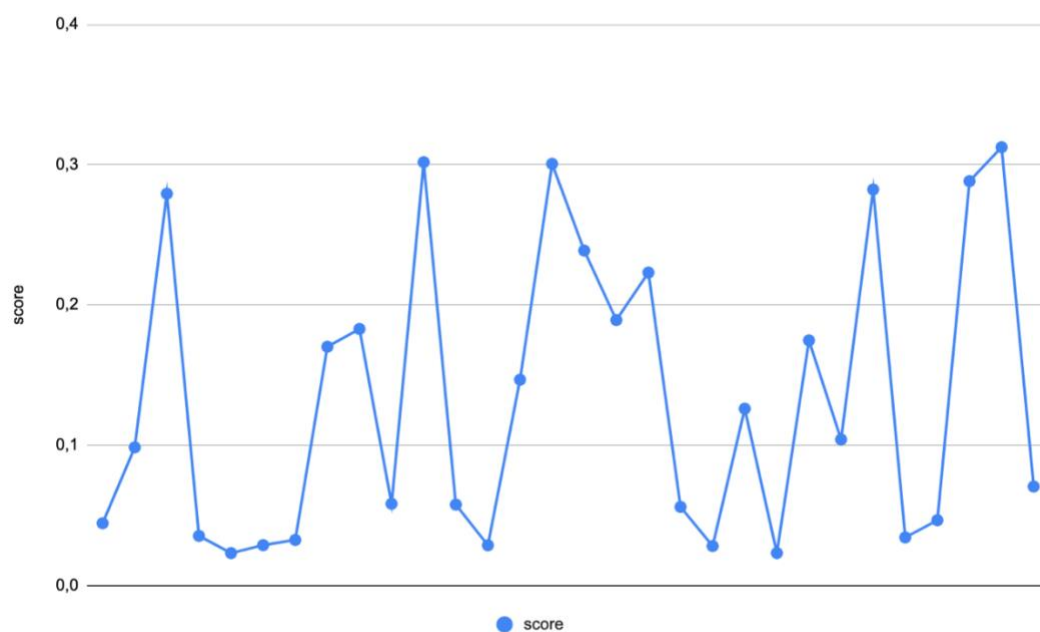


Σχήμα 8. Πραγματική και προβλεπόμενη παραγωγή για κάθε ένα από τα 30 τυχαία επιλεγμένα πεδία.

Επιπλέον, δημιουργήσαμε μια νέα μεταβλητή με την ονομασία "score" που δείχνει πόσο ακριβής είναι η πρόβλεψη. Η μεταβλητή "score" υπολογίζεται με βάση την ακόλουθη συνάρτηση:

$$\text{score} = \frac{\text{predicted yield} - \text{actual yield}}{\text{actual yield}}$$

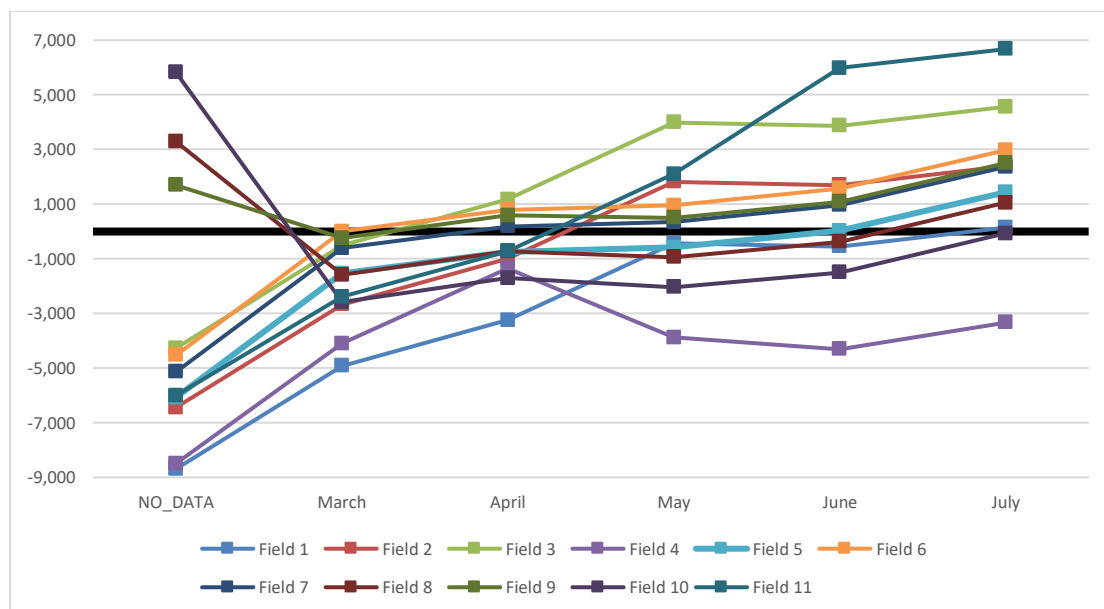
Στο Σχήμα 9, απεικονίζεται το σκορ κάθε από τα 30 πεδία.



### Σχήμα 9. Μεταβλητή σκορ για κάθε ένα από τα 30 τυχαία επιλεγμένα πεδία.

Σύμφωνα με τον ορισμό της μεταβλητής-σκορ, όσο πιο κοντά είναι το σκορ στο μηδέν, τόσο πιο ακριβής είναι η πρόβλεψη, καθώς ο αριθμητής είναι κοντά στο μηδέν, υποδηλώνοντας ότι η πρόβλεψη είναι σχεδόν πανομοιότυπη με την πραγματική τιμή. Το σκορ των επιλεγμένων πεδίων είναι μικρότερο από 0,4, πράγμα που σημαίνει ότι το μοντέλο είναι καλά εκπαιδευμένο και η πρόβλεψη είναι ακριβής.

Επιπλέον, δοκιμάσαμε το μοντέλο κατά τη διάρκεια μιας καλλιεργητικής περιόδου σε πραγματικές συνθήκες. Οι προβλέψεις πραγματοποιήθηκαν χρησιμοποιώντας δεδομένα για 11 διαφορετικά πεδία, για την καλλιέργεια του 2022, χρησιμοποιώντας δεδομένα από αγρότες τα οποία περιείχαν την υβριδική ποικιλία που καλλιεργήθηκε και την τοποθεσία του χωραφιού. Οι προβλέψεις έγιναν χωρίς να χρησιμοποιηθούν άλλα δεδομένα και ενημερώνονταν κάθε φορά που εισάγονταν νέα δεδομένα. Αυτή η διαδικασία πραγματοποιούνταν στο τέλος κάθε μήνα, ενημερώνοντας την πρόβλεψη. Στο Σχήμα 10 απεικονίζεται η διαφορά μεταξύ της προβλεπόμενης και της πραγματικής παραγωγής.



Σχήμα 10. Διαφορά μεταξύ προβλεπόμενης και πραγματικής σοδειάς κατά τη διάρκεια της καλλιέργειας για 11 πεδία κατά την περίοδο καλλιέργειας του 2022.

Όπως φαίνεται στο Σχήμα 10, όταν το μοντέλο προβλέπει χωρίς τη χρήση άλλων δεδομένων, οι προβλέψεις δεν είναι κοντά στην πραγματική παραγωγή. Ωστόσο, όταν εισάγονται νέα δεδομένα, οι προβλέψεις αρχίζουν να γίνονται πιο ακριβείς και η γραμμή πλησιάζει το κέντρο του γραφήματος (η πραγματική παραγωγή πλησιάζει την προβλεπόμενη παραγωγή). Μετά την προσθήκη των δεδομένων καιρού για τον Μάιο, το μοντέλο γίνεται πιο ακριβές και οι διαφορές στα περισσότερα χωράφια είναι κοντά στα 2.000 κιλά.

## Συζήτηση

Η πρόωρη και ακριβής πρόβλεψη της απόδοσης της παραγωγής αποτελεί ένα από τα σημαντικότερα προβλήματα στον γεωργικό τομέα. Η πρόβλεψη της παραγωγής είναι κρίσιμη για τους παραγωγούς από άποψη οικονομικής αξιολόγησης και για τον καθορισμό των σχεδίων διαχείρισής τους (Rashid et al., 2021; He et al., 2022). Ένας άλλος σημαντικός παράγοντας που πρέπει να ληφθεί υπόψη είναι οι απρόβλεπτες κλιματικές συνθήκες λόγω της αλλαγής του κλίματος, που μπορεί να επηρεάσουν την ανάπτυξη των φυτών και, κατά συνέπεια, την παραγωγή των καλλιεργειών με διάφορους τρόπους (Erda et al., 2005; Wang et al., 2018). Η επίδραση συγκεκριμένων κλιματικών συνθηκών κατά τη διάρκεια της

περιόδου καλλιέργειας τόνιστηκε στο Σχήμα 6, όπου το μοντέλο Ridge έδειξε πόσο σημαντικές είναι οι καιρικές συνθήκες για την τελική παραγωγή. Οι καιρικές συνθήκες μπορούν να αλλάξουν σε απόσταση λίγων χιλιομέτρων λόγω των μικροκλιματικών συνθηκών, ένα συνηθισμένο φαινόμενο στην Ελλάδα. Ως αποτέλεσμα, η πρόβλεψη που αφορά σε κάθε συγκεκριμένο χωράφι είναι σημαντική και η ενημέρωση της πρόβλεψης κατά την καλλιεργητική περίοδο, χρησιμοποιώντας τα πιο πρόσφατα δεδομένα καιρού, είναι μια λειτουργία που συμβάλλει στη δημιουργία ενός αξιόπιστου μοντέλου.

Μία από τις σημαντικότερες αποφάσεις στην ανάπτυξη ενός μοντέλου μηχανικής μάθησης είναι τα δεδομένα που θα χρησιμοποιηθούν. Σε αυτήν τη μελέτη, το μοντέλο μας εκπαιδεύτηκε χρησιμοποιώντας κλιματικά δεδομένα, καθώς είναι πιο εύκολα να αποκτηθούν και να χειριστούν. Δεδομένα κλίματος έχουν χρησιμοποιηθεί για την πρόβλεψη της παραγωγής σε άλλες καλλιέργειες, όπως σιταριού (Vagh and Xiao, 2012; Shakil Ahamed et al., 2015; Prasath et al., 2023), πατάτας (Kuradusenge et al., 2023; Prasath et al., 2023), καλαμποκιού (Crane-Droesch, 2018; Khaki and Wang, 2019; Khaki et al., 2020), σόγιας (Veenadhari et al., 2014), ρυζιού (Gandhi et al., 2017) και βαμβακιού (Prasad et al., 2021).

Ένας άλλος σημαντικός παράγοντας για την ανάπτυξη ενός μοντέλου μηχανικής μάθησης είναι ο αλγόριθμος που θα χρησιμοποιηθεί για την εκπαίδευση και την πρόβλεψη. Το μοντέλο που αναπτύχθηκε σε αυτήν τη μελέτη βασίστηκε στον αλγόριθμο Ridge, διότι είχε τις καλύτερες μετρικές σκορ για αυτό το σύνολο δεδομένων. Ο Ridge είναι μια τεχνική προσαρμογής που χρησιμοποιείται για δεδομένα με πολλαπλή συσχέτιση, η οποία μας βοηθά να κατανοήσουμε τα δεδομένα μας. Ένας άλλος αλγόριθμος που ευρέως χρησιμοποιείται για την πρόβλεψη της παραγωγής είναι το Random Forest, που αποτελείται από πολλά Decision Trees (Veenadhari et al., 2014; Kumar et al., 2020; Prasad et al., 2021; Kamath et al., 2021). Σε ορισμένες άλλες περιπτώσεις, τα δεδομένα δημιούργησαν ομάδες, πράγμα που σημαίνει ότι ένας αλγόριθμος ομαδοποίησης, όπως το K-Nearest, απέδωσε καλύτερα (Prasath et al., 2023). Ένας άλλος τρόπος εκπαίδευσης ενός μοντέλου είναι μέσω των Νευρωνικών Δικτύων, τα οποία προσαρμόζουν τη διανομή των βαρών με βάση την τιμή του σφάλματος κάθε εποχής. Τα Νευρωνικά Δίκτυα έχουν χρησιμοποιηθεί εκτενώς για την πρόβλεψη της παραγωγής καλλιεργειών (Crane-Droesch, 2018; Khaki and Wang, 2019; Khaki et al., 2020), ωστόσο, δεν απέδωσαν καλά σε αυτό το σύνολο δεδομένων. Τέλος, είναι συνηθισμένο να δοκιμάζονται διάφοροι αλγόριθμοι προκειμένου να βρεθεί αυτός που ταιριάζει καλύτερα σε κάθε σύνολο δεδομένων και στις ανάγκες μας, αντί να επιλέγεται ένας αλγόριθμος από προεπιλογή (Kuradusenge et al., 2023; Shakil Ahamed et al., 2015; Vagh and Xiao, 2012), όπως έγινε και σε αυτήν τη μελέτη. Σε αυτήν την περίπτωση, επιλέξαμε έναν αλγόριθμο που ταιριάζει καλά στο σύνολο δεδομένων αντί ενός που είναι πιο γρήγορος.

Για τον προσδιορισμό της ακρίβειας των προβλέψεων του μοντέλου, χρησιμοποιούνται οι καταγραφές των μετρικών του μοντέλου (RMSE,  $R^2$  και ακρίβεια). Το RMSE χρησιμοποιείται για να περιγράψει τη διαφορά μεταξύ των πραγματικών και των προβλεπόμενων τιμών. Σε μια μελέτη με πρόβλεψη παραγωγής σε διάφορες καλλιέργειες χρησιμοποιώντας Νευρωνικά Δίκτυα, το RMSE ήταν 5,274. Το μοντέλο Ridge που αναπτύχθηκε είχε χαμηλότερο RMSE (2,682), πράγμα που σημαίνει ότι κάνει πιο ακριβείς προβλέψεις (Shakil Ahamed et al., 2015). Σε μια άλλη μελέτη για την πρόβλεψη παραγωγής πατάτας και καλαμποκιού χρησιμοποιώντας τον αλγόριθμο Random Forest, η μετρική  $R^2$  ήταν 0,875 και 0,817 αντίστοιχα σε ένα σύνολο δεδομένων που αποτελείται από 1.921 εγγραφές για κάθε καλλιέργεια (Kuradusenge et al., 2023). Στη μελέτη μας, η μετρική  $R^2$  ήταν 0,144, το οποίο μπορεί να εξηγηθεί από το γεγονός ότι το σύνολο δεδομένων σε αυτήν τη μελέτη αποτελείται μόνο από 632 εγγραφές. Τέλος, όσον αφορά την ακρίβεια, η πρόβλεψη του μοντέλου Ridge έχει απόκλιση από τις πραγματικές τιμές που είναι μικρότερη από 0,4. Αυτή η μετρική χρησιμοποιείται για την ακρίβεια, δεδομένου ότι το αναπτυγμένο μοντέλο είναι ένα μοντέλο παλινδρόμησης και όχι ένας ταξινομητής.

Η πρόβλεψη της απόδοσης της παραγωγής κατά τη διάρκεια της καλλιεργητικής περιόδου επιτεύχθηκε σε προηγούμενες έρευνες χρησιμοποιώντας δεδομένα κλίματος και δεδομένα από δορυφόρους. Σε πρόσφατη έρευνα, αγρονομικές μετρήσεις και δεδομένα από αισθητήρες IoT, προερχόμενα από καλλιέργεια τομάτας σε θερμοκήπια, χρησιμοποιήθηκαν ως είσοδος σε ένα μοντέλο, με RMSE ίσο με 2.5974 (Lin et al., 2019). Σε μια παρόμοια έρευνα, δεδομένα από αισθητήρες IoT και του ιστορικά δεδομένα παραγωγής χρησιμοποιήθηκαν σε τρία διαφορετικά σύνολα δεδομένων. Το

αναπτυγμένο μοντέλο είχε RMSE που κυμαινόταν από 6,5 έως 10,45 (Gong et al., 2021). Σε μια άλλη έρευνα, εξήχθησαν δείκτες βλάστησης χρησιμοποιώντας δεδομένα από δορυφόρους σε μια προσπάθεια να προβλεφθεί η παραγωγή των ντοματών. Το τελικό μοντέλο είχε μια μέση τιμή σφάλματος 10% (Psiroukis et al., 2022). Σε μια έρευνα, χρησιμοποιώντας δεδομένα από δορυφόρους ως είσοδο, το μοντέλο προέβλεψε την παραγωγή σε ένα εύρος από -1 έως 1 τόνους ανά εκτάριο από την πραγματική (Dargra et al., 2023). Παρόλο που τα παραπάνω μοντέλα πρόγνωσης ήταν ακριβή στην πρόβλεψη της απόδοσης της παραγωγής τομάτας κατά τη διάρκεια της περιόδου καλλιέργειας, κανένα από αυτά δεν εκπαιδεύτηκε σε ανοικτό χωράφι με δεδομένα κλίματος. Τα δεδομένα κλίματος στις καλλιέργειες ανοικτού πεδίου είναι απρόβλεπτα, ωστόσο το μοντέλο που αναπτύχθηκε, ενσωμάτωσε μια ευρεία γκάμα δεδομένων καιρού, τα οποία μπορούν να εμπλουτιστούν περαιτέρω.

Τα μοντέλα μηχανικής μάθησης έχουν τη δυνατότητα να εντοπίζουν τον τρόπο με τον οποίο κάθε μεταβλητή στο σύνολο δεδομένων επηρεάζει τη μεταβλητή πρόβλεψης (Kasimatis et al., 2022). Η κύρια διαφορά μεταξύ της συσχέτισης που εξάγεται από τα μοντέλα μηχανικής μάθησης και της συσχέτισης που εξάγεται χρησιμοποιώντας παραδοσιακές στατιστικές μεθόδους είναι ότι τα μοντέλα μηχανικής μάθησης μπορούν να κατανοήσουν τα δεδομένα με ποικίλους τρόπους. Σε αυτήν τη μελέτη, ο συντελεστής που εξήχθη από το μοντέλο Ridge είχε ορισμένες σημαντικές διαφορές με τον συντελεστή που εξήχθη με παραδοσιακές μεθόδους. Μια έρευνα για τις καλλιέργειες στις ΗΠΑ, χρησιμοποιώντας δεδομένα κλίματος και εκπαίδευση με την χρήση γραμμικής παλινδρόμησης κατέληξε στο συμπέρασμα ότι η απόδοση των καλλιεργειών τομάτας είναι ανάλογη με τη μέγιστη θερμοκρασία του Απριλίου και του Ιουνίου (Lobell et al., 2007). Αυτό το αποτέλεσμα συμφωνεί με τα ευρήματά μας, όπου η μέγιστη θερμοκρασία του Απριλίου και του Ιουνίου είναι ανάλογη της τελικής απόδοσης. Σε μια άλλη μελέτη, η βροχόπτωση κατά τους καλοκαιρινούς μήνες ήταν η πιο σημαντική μεταβλητή σε ένα σύνολο δεδομένων που εκπαιδεύτηκε με Νευρωνικά Δίκτυα (Crane-Droesch, 2018). Η βροχόπτωση κατά τους καλοκαιρινούς μήνες ήταν επίσης σημαντική για το μοντέλο μας, δεδομένου ότι η βροχόπτωση του Ιουνίου και του Ιουλίου ήταν αντιστρόφως ανάλογη της απόδοσης για το μοντέλο. Σε μια έρευνα για την πρόβλεψη της απόδοσης της παραγωγής στην καλλιέργεια τομάτας, η θερμοκρασίας ήταν ανάλογη με την απόδοση της παραγωγής (Cho et al., 2021).

## Συμπεράσματα

Η ακριβής πρόβλεψη της απόδοσης της παραγωγής κατά τη διάρκεια ή ακόμα και πριν από την περίοδο καλλιέργειας είναι μια σημαντική πρόκληση για τους αγρότες και τις βιομηχανίες. Οι μετρήσεις που χρησιμοποιούνται στα αναπτυγμένα μοντέλα πρόγνωσης είναι εύκολο να αποκτηθούν και να επεξεργαστούν. Αυτό είναι σημαντικό, δεδομένου ότι δεν θα υπάρχει ανάγκη για την αντιμετώπιση ελλείψεων ή αταίριαστων δεδομένων. Σε αυτήν τη μελέτη, τα δεδομένα κλίματος και η υβριδική ποικιλία που χρησιμοποιήθηκε χρησιμοποιούνται ως είσοδο στο μοντέλο, το οποίο βασίζεται στον αλγόριθμο Ridge. Αυτός ο αλγόριθμος επιλέχθηκε λόγω των υψηλών σκορ που προέκυψαν στις μετρικές σφάλματος. Επιπλέον, το μοντέλο ήταν σε θέση να εντοπίσει τα σημαντικά δεδομένα στο σύνολο δεδομένων μας. Αυτό το χαρακτηριστικό είναι κρίσιμο, καθώς μας επιτρέπει να προβλέψουμε την βέλτιστη περίοδο συγκομιδής σε πραγματικό χρόνο, ακόμη και κατά τη διάρκεια της εξέλιξης της καλλιέργειας.

Στο μέλλον, θα ήταν σημαντικό να προσθέσουμε περισσότερα χαρακτηριστικά στο σύνολο δεδομένων μας (π.χ. δεδομένα από δορυφόρους, εικόνες RGB/ υπερφασματικές εικόνες από drones, δεδομένα ποιότητας, γεωργικές πρακτικές) προκειμένου να βελτιώσουμε την ακρίβεια των προβλέψεων, πέρα από την πραγματοποίηση προβλέψεων για τα χαρακτηριστικά ποιότητας της βιομηχανικής ντομάτας, όπως τα Total Soluble Solids ή τη λίπανση που πρέπει να εφαρμοστεί. Τέλος, είναι σημαντικό να εκπαιδεύσουμε το μοντέλο με δεδομένα που προέρχονται από διάφορες περιοχές της Ελλάδας, με διαφορετικές κλιματικές και εδαφικές συνθήκες καθώς και με διαφορετικά υβρίδια.

## Βιβλιογραφία



- Ara, I., Turner, L., Harrison, M.T., Monjardino, M., deVoil, P., Rodriguez, D., 2021. Application, adoption and opportunities for improving decision support systems in irrigated agriculture: A review. *Agric. Water Manag.* 257, 107161. <https://doi.org/10.1016/j.agwat.2021.107161>
- Belouz, K., Nourani, A., Zereg, S., Bencheikh, A., 2022. Prediction of greenhouse tomato yield using artificial neural networks combined with sensitivity analysis. *Sci. Hortic.* 293, 110666. <https://doi.org/10.1016/j.scienta.2021.110666>
- Boulila, W., Farah, I.R., Hussain, A., 2018. A novel decision support system for the interpretation of remote sensing big data. *Earth Sci. Inform.* 11, 31–45. <https://doi.org/10.1007/s12145-017-0313-7>
- Bouyoucos, G.J., 1962. Hydrometer Method Improved for Making Particle Size Analyses of Soils <sup>1</sup>. *Agron. J.* 54, 464–465. <https://doi.org/10.2134/agronj1962.00021962005400050028x>
- Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., Layton, R., Vanderplas, J., Joly, A., Holt, B., Varoquaux, G., 2013. API design for machine learning software: experiences from the scikit-learn project.
- Chawgien, K., Kiattisin, S., 2021. Machine learning techniques for classifying the sweetness of watermelon using acoustic signal and image processing. *Comput. Electron. Agric.* 181, 105938. <https://doi.org/10.1016/j.compag.2020.105938>
- Cho, W., Kim, S., Na, M., Na, I., 2021. Forecasting of Tomato Yields Using Attention-Based LSTM Network and ARMA Model. *Electronics* 10, 1576. <https://doi.org/10.3390/electronics10131576>
- Crane-Droesch, A., 2018. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environ. Res. Lett.* 13, 114003. <https://doi.org/10.1088/1748-9326/aae159>
- Darra, N., Espejo-Garcia, B., Kasimati, A., Kriezi, O., Psomiadis, E., Fountas, S., 2023. Can Satellites Predict Yield? Ensemble Machine Learning and Statistical Analysis of Sentinel-2 Imagery for Processing Tomato Yield Prediction. *Sensors* 23, 2586. <https://doi.org/10.3390/s23052586>
- Dehghanisani, H., Emami, S., Achite, M., Linh, N.T.T., Pham, Q.B., 2021. Estimating Yield and Water Productivity of Tomato Using a Novel Hybrid Approach. *Water* 13, 3615. <https://doi.org/10.3390/w13243615>
- Erda, L., Wei, X., Hui, J., Yinlong, X., Yue, L., Liping, B., Liyong, X., 2005. Climate change impacts on crop yield and quality with CO<sub>2</sub> fertilization in China. *Philos. Trans. R. Soc. B Biol. Sci.* 360, 2149–2154. <https://doi.org/10.1098/rstb.2005.1743>
- Gandhi, N., Armstrong, L.J., Nandawadekar, M., 2017. Application of data mining techniques for predicting rice crop yield in semi-arid climatic zone of India, in: 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR). Presented at the 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR), IEEE, Chennai, pp. 116–120. <https://doi.org/10.1109/TIAR.2017.8273697>
- Gandhi, N., Armstrong, L.J., Petkar, O., 2016. Proposed decision support system (DSS) for Indian rice crop yield prediction, in: 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR). Presented at the 2016 IEEE Technological Innovations in ICT for Agriculture and

Rural Development (TIAR), IEEE, Chennai, India, pp. 13–18. <https://doi.org/10.1109/TIAR.2016.7801205>

Gong, L., Yu, M., Cutsuridis, V., Kollias, S., Pearson, S., 2022. A Novel Model Fusion Approach for Greenhouse Crop Yield Prediction. *Horticulturae* 9, 5. <https://doi.org/10.3390/horticulturae9010005>

Gong, L., Yu, M., Jiang, S., Cutsuridis, V., Pearson, S., 2021. Deep Learning Based Prediction on Greenhouse Crop Yield Combined TCN and RNN. *Sensors* 21, 4537. <https://doi.org/10.3390/s21134537>

Gupta, A., Nahar, P., 2022. Classification and yield prediction in smart agriculture system using IoT. *J. Ambient Intell. Humaniz. Comput.* <https://doi.org/10.1007/s12652-021-03685-w>

He, L., Fang, W., Zhao, G., Wu, Z., Fu, L., Li, R., Majeed, Y., Dhupia, J., 2022. Fruit yield prediction and estimation in orchards: A state-of-the-art comprehensive review for both direct and indirect methods. *Comput. Electron. Agric.* 195, 106812. <https://doi.org/10.1016/j.compag.2022.106812>

Iraji, M.S., 2019. Comparison between soft computing methods for tomato quality grading using machine vision. *J. Food Meas. Charact.* 13, 1–15. <https://doi.org/10.1007/s11694-018-9913-2>

Kamath, P., Patil, P., S, Shrilatha, Sushma, S, Sowmya, 2021. Crop yield forecasting using data mining. *Glob. Transit. Proc.* 2, 402–407. <https://doi.org/10.1016/j.gltp.2021.08.008>

Kasimatis, C.-N., Psomakelis, E., Katsenios, N., Katsenios, G., Papatheodorou, M., Vlachakis, D., Apostolou, D., Efthimiadou, A., 2022. Implementation of a decision support system for prediction of the total soluble solids of industrial tomato using machine learning models. *Comput. Electron. Agric.* 193, 106688. <https://doi.org/10.1016/j.compag.2022.106688>

Katsenios, N., Andreou, V., Sparangis, P., Djordjevic, N., Giannoglou, M., Chanioti, S., Kasimatis, C.-N., Kakabouki, I., Leonidakis, D., Danalatos, N., Katsaros, G., Efthimiadou, A., 2022. Assessment of plant growth promoting bacteria strains on growth, yield and quality of sweet corn. *Sci. Rep.* 12, 11598. <https://doi.org/10.1038/s41598-022-16044-2>

Khaki, S., Pham, H., Wang, L., 2021. Simultaneous corn and soybean yield prediction from remote sensing data using deep transfer learning. *Sci. Rep.* 11, 11132. <https://doi.org/10.1038/s41598-021-89779-z>

Khaki, S., Wang, L., 2019. Crop Yield Prediction Using Deep Neural Networks. *Front. Plant Sci.* 10, 621. <https://doi.org/10.3389/fpls.2019.00621>

Khaki, S., Wang, L., Archontoulis, S.V., 2020. A CNN-RNN Framework for Crop Yield Prediction. *Front. Plant Sci.* 10, 1750. <https://doi.org/10.3389/fpls.2019.01750>

Kumar, Y.J.N., Spandana, V., Vaishnavi, V.S., Neha, K., Devi, V.G.R.R., 2020. Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector, in: 2020 5th International Conference on Communication and Electronics Systems (ICCES). Presented at the 2020 5th International Conference on Communication and Electronics Systems (ICCES), IEEE, Coimbatore, India, pp. 736–741. <https://doi.org/10.1109/ICCES48766.2020.9137868>

- Kuradusenge, M., Hitimana, E., Hanyurwimfura, D., Rukundo, P., Mtonga, K., Mukasine, A., Uwitonze, C., Ngabonziza, J., Uwamahoro, A., 2023. Crop Yield Prediction Using Machine Learning Models: Case of Irish Potato and Maize. *Agriculture* 13, 225. <https://doi.org/10.3390/agriculture13010225>
- Lagouvardos, K., Kotroni, V., Bezes, A., Koletsis, I., Kopania, T., Lykoudis, S., Mazarakis, N., Papagiannaki, K., Vougioukas, S., 2017. The automatic weather stations NOANN network of the National Observatory of Athens: operation and database. *Geosci. Data J.* 4, 4–16. <https://doi.org/10.1002/gdj3.44>
- Lázaro, E., Makowski, D., Vicent, A., 2021. Decision support systems halve fungicide use compared to calendar-based strategies without increasing disease risk. *Commun. Earth Environ.* 2, 224. <https://doi.org/10.1038/s43247-021-00291-8>
- Leoni, C., 2003. FOCUS ON LYCOPENE. *Acta Hortic.* 357–363. <https://doi.org/10.17660/ActaHortic.2003.613.58>
- Li, N., Wu, X., Zhuang, W., Xia, L., Chen, Y., Wu, C., Rao, Z., Du, L., Zhao, R., Yi, M., Wan, Q., Zhou, Y., 2021. Tomato and lycopene and multiple health outcomes: Umbrella review. *Food Chem.* 343, 128396. <https://doi.org/10.1016/j.foodchem.2020.128396>
- Lin, D., Wei, R., Xu, L., 2019. An Integrated Yield Prediction Model for Greenhouse Tomato. *Agronomy* 9, 873. <https://doi.org/10.3390/agronomy9120873>
- Liu, J., 2018. Tomato Yield Estimation Based on Object Detection. *J. Adv. Comput. Intell. Intell. Inform.* 22, 1120–1125. <https://doi.org/10.20965/jaciii.2018.p1120>
- Lobell, D.B., Cahill, K.N., Field, C.B., 2007. Historical effects of temperature and precipitation on California crop yields. *Clim. Change* 81, 187–203. <https://doi.org/10.1007/s10584-006-9141-3>
- M. Koller, S. K. Upadhyaya, 2005. Prediction of Processing Tomato Yield Using a Crop Growth Model and Remotely Sensed Aerial Images. *Trans. ASAE* 48, 2335–2341. <https://doi.org/10.13031/2013.20072>
- Maheswari, P., Raja, P., Hoang, V.T., 2022. Intelligent yield estimation for tomato crop using SegNet with VGG19 architecture. *Sci. Rep.* 12, 13601. <https://doi.org/10.1038/s41598-022-17840-6>
- Makris, A., Kontopoulos, I., Psomakelis, E., Xyalis, S.N., Theodoropoulos, T., Tserpes, K., 2022. Performance Analysis of Storage Systems in Edge Computing Infrastructures. *Appl. Sci.* 12, 8923. <https://doi.org/10.3390/app12178923>
- Malviya, A., Solanki, Prof.D.S., 2022. Crop Yield Prediction Using Deep Neural Networks. *Int. J. Res. Appl. Sci. Eng. Technol.* 10, 657–665. <https://doi.org/10.22214/ijraset.2022.46226>
- Nigam, A., Garg, S., Agrawal, A., Agrawal, P., 2019. Crop Yield Prediction Using Machine Learning Algorithms, in: 2019 Fifth International Conference on Image Information Processing (ICIIP). Presented at the 2019 Fifth International Conference on Image Information Processing (ICIIP), IEEE, Shimla, India, pp. 125–130. <https://doi.org/10.1109/ICIIP47207.2019.8985951>
- Papageorgiou, E.I., Markinos, A.T., Gemtos, T.A., 2011. Fuzzy cognitive map based approach for predicting yield in cotton crop production as a basis for decision support system in precision

agriculture application. *Appl. Soft Comput.* 11, 3643–3657. <https://doi.org/10.1016/j.asoc.2011.01.036>

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.

Prabakaran, G., Vaithyanathan, D., Ganesan, M., 2018. Fuzzy decision support system for improving the crop productivity and efficient use of fertilizers. *Comput. Electron. Agric.* 150, 88–97. <https://doi.org/10.1016/j.compag.2018.03.030>

Prasad, N.R., Patel, N.R., Danodia, A., 2021. Crop yield prediction in cotton for regional level using random forest approach. *Spat. Inf. Res.* 29, 195–206. <https://doi.org/10.1007/s41324-020-00346-6>

Prasath, N., Sreemathy, J., Krishnaraj, N., Vigneshwaran, P., 2023. Analysis of Crop Yield Prediction Using Random Forest Regression Model, in: So-In, C., Londhe, N.D., Bhatt, N., Kitsing, M. (Eds.), *Information Systems for Intelligent Systems, Smart Innovation, Systems and Technologies*. Springer Nature Singapore, Singapore, pp. 239–249. [https://doi.org/10.1007/978-981-19-7447-2\\_22](https://doi.org/10.1007/978-981-19-7447-2_22)

Psiroukis, V., Darra, N., Kasimati, A., Trojacek, P., Hasanli, G., Fountas, S., 2022. Development of a Multi-Scale Tomato Yield Prediction Model in Azerbaijan Using Spectral Indices from Sentinel-2 Imagery. *Remote Sens.* 14, 4202. <https://doi.org/10.3390/rs14174202>

Psomakelis, E., Aisopos, F., Litke, A., Tserpes, K., Kardara, M., Campo, P.M., 2016. Big IoT and Social Networking Data for Smart Cities - Algorithmic Improvements on Big Data Analysis in the Context of RADICAL City Applications:, in: *Proceedings of the 6th International Conference on Cloud Computing and Services Science. Presented at the Workshop on Towards Convergence of Big Data, SQL, NoSQL, NewSQL, Data streaming/CEP, OLTP and OLAP, SCITEPRESS - Science and Technology Publications*, Rome, Italy, pp. 396–405. <https://doi.org/10.5220/0005934503960405>

Psomakelis, E., Nikolakopoulos, A., Marinakis, A., Psychas, A., Moulos, V., Varvarigou, T., Christou, A., 2020a. A Scalable and Semantic Data as a Service Marketplace for Enhancing Cloud-Based Applications. *Future Internet* 12, 77. <https://doi.org/10.3390/fi12050077>

Psomakelis, E., Tserpes, K., Zissis, D., Anagnostopoulos, D., Varvarigou, T., 2020b. Context agnostic trajectory prediction based on  $\lambda$ -architecture. *Future Gener. Comput. Syst.* 110, 531–539. <https://doi.org/10.1016/j.future.2019.09.046>

Rashid, M., Bari, B.S., Yusup, Y., Kamaruddin, M.A., Khan, N., 2021. A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches With Special Emphasis on Palm Oil Yield Prediction. *IEEE Access* 9, 63406–63439. <https://doi.org/10.1109/ACCESS.2021.3075159>

Rupnik, R., Kukar, M., Vračar, P., Košir, D., Pevec, D., Bosnić, Z., 2019. AgroDSS: A decision support system for agriculture and farming. *Comput Electron Agric* 161, 260–271. <https://doi.org/10.1016/j.compag.2018.04.001>.

Salazar, R., Dannehl, D., Schmidt, U., López, I., Rojano, A., 2017. A dynamic artificial neural network for tomato yield prediction. *Acta Hort.* 83–90. <https://doi.org/10.17660/ActaHortic.2017.1154.11>

- Salazar, R., López, I., Rojano, A., Schmidt, U., Dannehl, D., 2015. Tomato yield prediction in a semi-closed greenhouse. *Acta Hortic.* 263–270. <https://doi.org/10.17660/ActaHortic.2015.1107.36>
- Shakil Ahamed, A.T.M., Mahmood, N.T., Hossain, N., Kabir, M.T., Das, K., Rahman, F., Rahman, R.M., 2015. Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh, in: 2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). Presented at the 2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), IEEE, Takamatsu, pp. 1–6. <https://doi.org/10.1109/SNPD.2015.7176185>
- Small, I.M., Joseph, L., Fry, W.E., 2015. Development and implementation of the BlightPro decision support system for potato and tomato late blight management. *Comput. Electron. Agric.* 115, 57–65. <https://doi.org/10.1016/j.compag.2015.05.010>
- Tatsumi, K., Igarashi, N., Mengxue, X., 2021. Prediction of plant-level tomato biomass and yield using machine learning with unmanned aerial vehicle imagery. *Plant Methods* 17, 77. <https://doi.org/10.1186/s13007-021-00761-2>
- Vagh, Y., Xiao, J., 2012. Mining temperature profile data for shire-level crop yield prediction, in: 2012 International Conference on Machine Learning and Cybernetics. Presented at the 2012 International Conference on Machine Learning and Cybernetics (ICMLC), IEEE, Xian, Shaanxi, China, pp. 77–83. <https://doi.org/10.1109/ICMLC.2012.6358890>
- Veenadhari, S., Misra, B., Singh, C., 2014. Machine learning approach for forecasting crop yield based on climatic parameters, in: 2014 International Conference on Computer Communication and Informatics. Presented at the 2014 International Conference on Computer Communication and Informatics (ICCCI), IEEE, Coimbatore, India, pp. 1–5. <https://doi.org/10.1109/ICCCI.2014.6921718>
- Wang, J., Vanga, S., Saxena, R., Orsat, V., Raghavan, V., 2018. Effect of Climate Change on the Yield of Cereal Crops: A Review. *Climate* 6, 41. <https://doi.org/10.3390/cli6020041>
- Wang, X., Vladislav, Z., Viktor, O., Wu, Z., Zhao, M., 2022. Online recognition and yield estimation of tomato in plant factory based on YOLOv3. *Sci. Rep.* 12, 8686. <https://doi.org/10.1038/s41598-022-12732-1>
- Wang, Y., Xiao, R., Yin, Y., Liu, T., 2021. Prediction of Tomato Yield in Chinese-Style Solar Greenhouses Based on Wavelet Neural Networks and Genetic Algorithms. *Information* 12, 336. <https://doi.org/10.3390/info12080336>
- Zimmermann, M., Fischer, M., 2020. Impact assessment of water and nutrient reuse in hydroponic systems using Bayesian Belief Networks. *J. Water Reuse Desalination* 10, 431–442. <https://doi.org/10.2166/wrd.2020.026>

